DOCUMENT RESUME

ED 161 950 TM 007 991

AUTHOR	Anderson, Beverly; And Others
TITLE	Report of the Michigan Educational Assessment Program's External Advisory Panel on Evaluation.
SPONS AGENCY	Michigan State Dept. of Education, Lansing.; National Assessment of Educational Progress, Ann Arbor, Mich.
PUB DATE	Jul 77
NOTE	72p.; For related document, see TM 007 316
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	Check Lists; *Educational Assessment; Educational Objectives; *Evaluation Criteria; Evaluation Methods; *Measurement Goals; Program Administration; Program Costs; Program Effectiveness; *Program Evaluation; *State Programs; *Testing Programs
IDENTIFIERS	*Michigan Educational Assessment Program; Shepard Evaluation Model

ABSTRACT

This report is an evaluation of the Michigan Educational Assessment Program (MEAP) done by an external advisory panel. Panelists were selected in terms of availability and complementary skills. The program was evaluated by the Shepard method which consists of a checklist grouped into the areas of goals and purposes, technical aspects, management, intended and unintended effects, and costs. Each question on the checklist was given, followed by identification of major documents or other evidence obtained and the panel's response. In some instances, tentative judgments were made or recommendations for how to gather additional evidence or suggestions for improving a given area. Section II suggest possible ways MEAP could respond to the panel's findings. Appendices include: panel selection and biases, memorandum of agreement, Shepard checklist and resulting questions, a 4-page review of the Michigan Educational Assessment Program (March 1977), a list of the 29 people interviewed, and a 103-item bibliography. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Malcom Katz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED161950

IM007 991

Report of the Michigan Educational Assessment Program's External Advisory Panel on Evaluation

Under a Contract with
Michigan Department of Education and
National Assessment of Educational Progress

July 1977

REPORT OF THE MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM'S EXTERNAL ADVISORY PANEL ON EVALUATION

by

Beverly Anderson
Northwest Regional Educational Laboratory
Portland, Oregon

William Cooley
Learning Research and Development Center
University of Pittsburgh

Al Holliday
Educational Communication Center
Camp Hill, Pennsylvania

William Mosley
Special Education Department
Indiana University

Augustus B. Turnbull III
Public Administration Department
Florida State University

Under a Contract with

Michigan Department of Education and
National Assessment of Educational Progress

July 1977

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	USE OF THE REPORT	5
III.	GOALS AND PURPOSES	9
IV.	TECHNICAL ASPECTS	15
V.	MANAGEMENT	25
VI.	INTENDED AND UNINTENDED EFFECTS	33
VII.	COSTS	37
VIII.	SUMMARY	39
APPENDIX A	Panel Selection and Biases	41
APPENDIX B	Memorandum of Agreement	51
APPENDIX C	Shepard Checklist and Resulting Questions	57
APPENDIX D	A Review of the Michigan Educational Assessment Program, March 1977	69
APPENDIX E	Documents Reviewed and People Interviewed	75

I. INTRODUCTION

In March 1977 representatives of the Michigan Educational Assessment Program (MEAP) and the National Assessment of Educational Progress (NAEP) met with Loretta Shepard, author of a method of evaluating assessments,[1] to make plans for an evaluation of MEAP. NAEP was interested in having the Shepard evaluation method used to evaluate a statewide assessment program and MEAP was interested in having such an evaluation conducted. Thus the two organizations chose to jointly sponsor the work of an external panel which would apply the Shepard evaluation method to MEAP.

At the March meeting a five-member panel was identified according to guidelines presented in the Shepard evaluation method.[2] Each panel member was to devote five days to the task (with the team leader having an additional five days). The Shepard checklist, which provides nearly 50 topics to investigate when doing an evaluation of an assessment program, was to guide the work of the panel. The topics in the Shepard checklist are grouped into five areas: (a) goals and purposes, (b) technical aspects, (c) management, (d) intended and unintended effects, and (e) costs.

On April 27, 1977, Judith Moyer of MEAP, Jack Schmidt of NAEP and Loretta Shepard met with the panel in Chicago to clarify the panel's responsibilities and operating procedures. Given the desire to have the panel's work completed by June, it was necessary to limit the scope

---

[1] Shepard, L. A. A method of evaluating assessment. Paper presented at the Sixth Annual Conference on Large Scale Assessment, Boulder, Colorado, June 1976.

[2] Appendix A contains a statement of how panelists were selected, as well as a summary of a questionnaire which panelists completed to determine the nature of their biases on testing prior to reviewing MEAP.

or procedures of the study. The panel, jointly with MEAP, agreed that all five areas of the checklist needed to be considered. And, given the time constraints, it was agreed that the most useful information could be obtained if the panel focused on *the extent to which MEAP has and is gathering useful information to serve as the basis for decisions in each of the above five areas.* The panel was not commissioned to generate new data, e.g., about the technical quality of the tests or the program effects, but rather to judge whether or not MEAP has adequate information available to make decisions regarding the topics presented in the checklist. Thus, the panel did not conduct an evaluation of MEAP.

It was further decided that the panel's approach would be to review existing documents and interview key decision makers and MEAP staff to see if adequate information exists. Appendix B contains the memorandum of agreement which resulted from the April 27 meeting.

At the April meeting the panel also reviewed the Shepard checklist and developed specific questions to guide their review. Appendix C contains the Shepard checklist, with the panel members' additions in italics and their translation of the checklist categories into specific questions. It was necessary to continually keep in mind that the primary task had been defined as that of determining if sufficient evidence existed regarding the topics. Each question was therefore phrased in terms of "Does sufficient evidence exist to judge the adequacy of. . ."

Using the list of questions developed by the panel, the MEAP staff sent each panelist relevant documents to review during May. After reviewing documents, each panelist identified people (by name or position) whom they wished to interview. The MEAP contact person made arrangements for these interviews. The panel members did not attempt to interview people beyond MEAP staff, legislators, news media representatives,

Michigan Education Association (MEA) personnel and key Michigan Department of Education (MDE) decision makers, since they were focusing only on the extent to which evidence now is available to MEAP staff to have an evaluation of the program done.

From June 1 to 3 panelists conducted interviews in Lansing with MEAP staff, legislators, news media personnel, MEA staff and key MDE decision makers and wrote an initial draft of their report. On June 16 the team leader made a presentation at the Seventh Annual Conference on Large Scale Assessment on the panel's response to the use of the Shepard method and some findings of the panel.[3]

The present report represents the completion of the panel's work under the agreement among the panel, MEAP and NAEP. The content has been reviewed and approved by all panel members. A draft version was also reviewed by the MEAP Supervisor to ensure factual accuracy. He was not asked to respond to the specific recommendations made by the panel; the panel made the final decisions about the report's content and are therefore solely responsible for it. Judgments expressed in the report are those of the panel, based on the evidence they examined.

Structure of the Report

Section II of the report presents suggestions on how to use the report. As stated earlier, the report does not represent an evaluation of MEAP but rather is intended to assist the staff as they improve the continuous evaluation of MEAP.

---

[3] Anderson, B. L. Is Michigan assessment as--a. good, b. bad--as everyone says? A tryout of the Shepard Evaluation Model. Paper presented at the Seventh Annual Conference on Large Scale Assessment, Boulder, Colorado, June 1977.

Sections III through VII of the report correspond to the five sections of the Shepard checklist--

- goals and purposes
- technical aspects
- management
- intended and unintended effects
- cost.

Each question addressed by the panel relative to that program aspect is given, followed by identification of the major documents or other evidence the panel obtained and the panel's response to the question. Note that the panel's response often is not restricted solely to stating whether or not adequate evidence exists to judge the particular aspect of the program (the panel's primary focus). In some instances, the panel made tentative judgments based on the evidence, recommendations for how to gather additional evidence, and at times, suggestions for how to improve a given area. *It is essential to realize that such statements which go beyond the primary task are not intended to be a comprehensive evaluation of that aspect of the program. They are only initial reactions which may be helpful in future planning and the conduct of an evaluation.*

Since this report is intended primarily for MEAP staff, a lengthy description of the program will not be included. However, Appendix D, prepared by a MEAP staff member, describes MEAP's history, present program and possible future directions.

II. USE OF THE REPORT

A major issue concerning any evaluation or evaluation related effort is how the report is to be used. This report is intended to serve as part of a planning process for future evaluation. The information presented herein is intended to guide MEAP staff members as they undertake the collection of additional information.

The panel emphasizes that MEAP should not necessarily plan to obtain evidence on all areas covered in the report; some priorities must be set. The cost of obtaining documentation may far outweigh the benefits in certain areas. MEAP may also have to choose between going back to document past work or planning for documentation of similar work in the future. Several other options exist. The panel has identified the following as possible responses which MEAP could make to the panel's comments:

1. Where sufficient evidence now exists, MEAP could hire an external person, with expertise in the particular area being judged, to review the evidence and make judgments of its quality and/or recommendations for change.

2. Where sufficient evidence exists, MEAP could have internal staff members review documents and make judgments of quality and/or recommendations for change.

3. Where insufficient evidence exists, MEAP could prepare documentation of past occurrences so judgments can be made about a given aspect of the program.

4. Where insufficient evidence exists, MEAP could choose not to document past work but rather make plans to write future documents on the topic so adequate judgments can be made either (a) internally or (b) externally.

5. MEAP could have internal staff gather additional data on the issue which (a) they or (b) internal staff can use to make judgments.

6. MEAP could have an external group gather additional data on the issue which (a) they or (b) internal staff can use to make judgments.

7. If it appears that something was not documented because it never occurred, MEAP could institute a planning process relative to that aspect of the program. Such planning may be done by an (a) internal and/or (b) external group.

8. MEAP could choose to do nothing at the present time.

In making choices among the above eight alternatives (or others the staff may suggest), the following questions should be considered:

1. What decisions will be affected by the data?

2. What time limitations exist?

3. What financial limitations exist?

4. Is internal staff available for such efforts?

5. Are appropriate external consultants available for such efforts?

6. How frequently are various kinds of data needed?

The following overall perspective of what the panel considers to be key decisions which MEAP or other Department staff need to make may be helpful in the critical task of setting priorities on evaluation activities.[4] The panel considers the following decisions to be imperative in guiding future planning and evaluation:

1. Which MEAP goals are most important and should serve as the basis for designing the assessment content and procedures?

2. How and to what extent should assessment be expanded?[5]

3. Which content objectives should be tested within reading and math?

---

[4] When developing the list of key decisions it is important to identify the people who will be making the decisions. Such people need to be involved in planning to ensure that they will find the data useful in decision making.

[5] It is critical to establish the quality of what already exists and its value to various audiences before expansion occurs.

4. Should all students be tested on the same objectives?

5. What staff development is needed?

6. What changes should be made in reporting and dissemination?

7. What changes are needed in the item writing and review process to eliminate bias in the test items against the aged, handicapped, either sex, or any cultural groups?

8. Which decision making groups should receive priority when compromises must be made--state, district, school, classroom, student?

One further comment must be made about the information in Sections III through VII. The panel recognizes that they may have missed some documents; they have made statements in terms of the evidence they were given. Corrections should be made and new evidence introduced where it may make differences in future planning.

III. GOALS AND PURPOSES

The purpose of this section is to consider the existing evidence regarding the goals and purposes of MEAP. Two recent statements of MEAP objectives exist. A July 1975 document, "Objectives and Procedures" (prepared by MEAP), presents the following objectives:

1. The first objective of the MEAP is to provide the State Board of Education, the Executive Office, the Legislature and citizens with data describing the levels of basic skills attainment and other relevant descriptive data about each of Michigan's schools and school districts.

2. The second objective of the MEAP is to provide local educators with specific information about the levels of educational attainment of students in the basic skills and other relevant descriptive data for their own schools and school districts.

3. The third objective of the Michigan Educational Assessment Program is to provide information regarding the progress of the entire Michigan educational system and the progress of its school districts and schools over a period of years to the State Board of Education, the Executive Office, the Legislature and citizens.

A May 1976 document, "Long Range Plan of the Michigan Educational Assessment Program" (prepared by the MEAP), identifies these objectives:

1. To provide statewide information describing the current status of pupil achievement of minimum levels of basic skills and other common learnings and the current status of human and financial resources.

2. To provide statewide information describing what progress is being made over time in pupil achievement of minimum levels of basic skills and other common learnings and what changes are occurring in the status of human and financial resources.

3. To provide statewide information that can help to identify areas of success and areas of need that can be used as a partial basis for allocation of state level human, financial and material resources designed to help ensure pupil achievement of minimum levels of basic skills and other common learnings.

4. To provide a partial base of information to local school districts to assist them in:

   a. describing the current status of pupil achievement of minimum levels of basic skills and other common learnings;
   b. describing the progress being made over time in pupil achievement of these minimum levels;
   c. identifying district, school classroom, and individual pupil areas of success and areas of need for purposes of further anlayzing their own programs and, when appropriate, reallocating their own human, financial and material resources in ways designed to help ensure pupil achievement of these minimum levels; and
   d. reporting to parents on the status and progress of their child's attainments in school.

5. To provide technical assistance and support to local districts in the development and implementation of comprehensive local needs assessment programs.

6. To establish standards and instrumentation for individual pupil assessment of the achievement of minimum levels of basic skills and other common learnings.

7. To promote optimum utility of assessment data by actively promoting the dissemination and utilization of Michigan assessment results to all concerned audiences, including parents.

8. To provide the State Board of Education with an annual systematic audit of the policies and procedures of the Michigan Educational Assessment Program, including input from the REAS Advisory Council and other appropriate advisory groups.

Below are the questions which the panel investigated relative to goals and purposes. Each question is followed by identification of major sources of evidence and the panel's response. The first questions presented tend to be those for which the most evidence existed, whereas those at the end of the section are the ones for which little or no evidence was seen.

Question 1:	Does sufficient evidence exist to determine if goals have been established?
Major Evidence:[6]	Documents - 22, 23
Panel Response:	Adequate evidence exists to determine if goals have been established. Goals are clearly stated. The goals in the long range plan go beyond those of the 1976 document and also go beyond use of assessment data into operation of the assessment program.

Question 2:	Does sufficient evidence exist to judge the adequacy of consideration given to the following kinds of goals: pupil diagnosis; pupil certification; program evaluation; improvement; selection; resource allocation; accountability; research; informing the public?
Major Evidence:	Documents - 4, 5, 7, 15, 22-26 Interviews - 2, 7, 21, 23
Panel Response:	Adequate evidence exists. Stated goals show evidence of consideration of pupil diagnosis, resource allocation, accountability, informing the public, research related goals. Interviews and documents indicate decisions have been made to exclude pupil certification and program evaluation as goals although such decisions may be reconsidered.

Question 3:	Does sufficient evidence exist to judge the adequacy of the process used to establish goals?
Major Evidence:	Documents - 4, 5, 22, 23 Interviews - 6, 17, 21, 23
Panel Response:	Inadequate evidence exists. A statement of when, who and how goals were established is needed.

[6] Codes used here refer to the documents and interviews listed in Appendix E.

Question 4: Does sufficient evidence exist to judge the adequacy of the existing process used to modify goals?

Major Evidence: Documents - 4, 5, 22, 23

Interviews - 6, 17, 21, 23

Panel Response: The lack of evidence seen by the panel suggests that such a process does not exist. If accurate, plans should be made to develop a process for goal modification which states how, when and the basis for such modification.

Question 5: Does sufficient evidence exist to judge the clarity of understanding among affected groups concerning what the goals are?

Major Evidence: Documents - 22, 23

Interviews - 7, 10, 18, 21

Panel Response: The panel found that different groups (MEAP staff, legislators or aides, MEA, Superintendent of Public Instruction) emphasize different objectives, some of which are contained in MEAP's official documents and some of which are not. It is the panel's opinion that decisions concerning which goals to retain, emphasize, eliminate or add need to be made on the basis of objectively gathered information. One possible procedure for modifying goals and clarifying understanding among groups would be to obtain statements from legislators, the Superintendent of Public Instruction, MEAP staff, principals, local district test coordinators, teachers, educational associations and the public concerning what they perceive the goals are at the present time. To facilitate modification of goals, it may be useful also to ask such people what they believe goals <u>should</u> be. When goals are modified, it is also important to determine who has the right to establish the legitimate goals of MEAP--the legislature, the Superintendent, MEAP staff or others.

Question 6: Does sufficient evidence exist to judge the adequacy of the importance, uniqueness, feasibility and appropriateness of the scope of the goals?

Panel Response: The extent to which goals are collectively feasible, unqiue and appropriate cannot be judged on existing evidence. Extensive information is needed on the extent to which MEAP is, in fact, following stated goals and whether MEAP's efforts are having their intended effects.

Despite the official statement of objectives, more information needs to be gathered to serve as a basis for clarifying the congruence among MEAP's objectives, its effects and the way tests are constructed and sampling is done. The differences in emphasis imply different approaches to test construction, sampling and valuing of effects. See Section VI for data gathering suggestions.

IV. TECHNICAL ASPECTS

The panel considered 18 questions related to the technical aspects of MEAP. Many documents related to technical aspects have been prepared by advisory groups, contractors and MEAP staff, indicating that considerable attention has been directed toward the technical issues of assessment. Below are the questions which the panel investigated relative to technical aspects of MEAP. Each question is followed by identification of major sources of evidence and the panel's response.

Question 1: Does sufficient evidence exist to judge the adequacy of the basis for selecting norm referenced or objective referenced tests?

Major Evidence: Documents - 16, 22, 35

Interviews - 2, 7, 16, 21, 23

Panel Response: Although little documentation was available on the basis for the change, interviews indicated that the change was made because of the desire to have information related to specific skills. It is not apparent whose opinions--teachers, test experts, content specialists--were considered. The panel does not consider it a high priority to document now how this past decision was made; rather, attention should be directed towards determining the present usefulness of the objective referenced tests.

Question 2: Does sufficient evidence exist to judge the consistency of goals to the choice of norm referenced or objective referenced tests?

Major Evidence: Documents - 2, 16, 27

Interviews - 2, 7, 16, 21

Panel Response: No evidence was presented to show that the choice was carefully looked at in terms of the consistency with the goals. There is a need to determine which goals are most consistent with each test type.

Question 3:	Does sufficient evidence exist to judge the adequacy of the process used to select content areas for testing?
Major Evidence:	Documents - 1, 7, 15, 16, 17, 24, 35 Interviews - 21, 24, 26, 28
Panel Response:	No documentation was seen of how and by whom test content areas were selected. Documents only state that selection occurred.

Question 4:	Does sufficient evidence exist to judge the adequacy of the conduct and use of item analysis?
Major Evidence:	Documents - 16, 35, 55, 61 Interviews - 26
Panel Response:	Item analyses are well documented. Documents and interviews indicate that item analyses are being appropriately used in test revision.

Question 5:	Does sufficient evidence exist to judge the adequacy of the test administration process?
Major Evidence:	Documents - 15, 22, 65 Interviews - 16, 21
Panel Response:	The documents clearly state the responsibilities of administrators and teachers in the test administration process. Feedback has been gathered on administrative difficulties.

Question 6:	Does sufficient evidence exist to judge the adequacy of the match of test items to content objectives?
Major Evidence:	Documents - 1, 2, 15, 17 Interviews - 2, 16, 21, 24, 26, 28
Panel Response:	The documents reviewed do not present the criteria used to determine if a given item did indeed match the objective for which it was written. Interviews indicate that the opinions of experts such as Michigan Reading Association (MRA) and Michigan Council of

Teachers of Mathematics (MCTM) representatives, serve as the basis for determining congruence of items and objectives.

Question 7: Does sufficient evidence exist to judge the adequacy of the method of handling missing data?

Major Evidence: Documents - 15, 22, 65

Interviews - 16, 26

Panel Response: The documents examined do not specify the precise procedures used for handling missing data.

Question 8: Does sufficient evidence exist to judge the adequacy of the longitudinal analyses?

Major Evidence: Documents - 40, 103

Panel Response: The evidence indicates that longitudinal data analyses are being appropriately conducted and presented in reports. No evidence exists, however, on how or the extent to which longitudinal data are being used.

Question 9: Does sufficient evidence exist to judge the usefulness of reporting measures (e.g., human resources)?

Major Evidence: Documents - 40, 103

Interviews - 3, 4, 10, 19, 27

Panel Response: Evidence is available to judge the quality of the reporting measures data. The panel did not see evidence to indicate (1) a recent review of the appropriateness of the set of reporting measures now being used, nor (2) the extent to which the present reporting measures are useful.

Question 10: Does sufficient evidence exist to judge how credible audiences consider assessment data to be?

Major Evidence: Documents - 7, 41, 48, 50

Interviews - 7, 12, 16, 21

Panel Response: The credibility of data has been considered only with regard to the appropriateness of objectives selected and the match of items to objectives. No other evidence was seen on credibility of data (e.g., longitudinal, relation to economic factors).

Question 11: Does sufficient evidence exist to judge the adequacy of the procedures used to field test items and tests?

Major Evidence: Documents - 2, 17, 22, 35, 62, 65

Interviews - 26

Panel Response: The field testing process appears satisfactory. Care is being taken to balance practical constraints with the need for representative districts.

Question 12: Does sufficient evidence exist to judge the adequacy of the reporting procedures (e.g., appropriateness of data analyses, use of different reports for different audiences, appropriateness of interpretation)?

Major Evidence: Documents - 10, 15, 18, 20, 22, 26, 33, 34, 49, 64, 103

Interviews - 2, 3, 4, 10, 16, 19, 21, 26, 27

Panel Response: Existing documents to identify the reporting procedures used. General data exist to indicate the usefulness of various reports. More specific information is needed from audiences concerning how reports are used and what changes are needed to make them most appropriate to their audiences. More information is needed on the effects of MEAP and its testing to determine if reporting procedures mesh with the desired effects. Insufficient evidence now exists to say that the materials MEAP uses for interpretation are adequate for users such as news media, boards of education and parents. Discussions with several news media representatives indicate it would be beneficial to them to receive assistance from MEAP in interpreting both

statewide and local results. A brief publication for the media on the topic "How to Interpret Assessment Results" in the language of laypeople would probably prevent reporters from making gross interpretaton errors and in the long run would orient readers as to what assessment is and is not. An advisory committee of a few reporters could meet with MEAP staff to draft such a statement. In light of the great turnover of reporters, the how-tos should be sent with each year's releases and data reports. Copies could be sent to superintendents for use with local media.

The difficulty MEAP has had in expanding the scope of assessment may well reflect the judgment of decision makers that current reports are not useful--so why expand? Legislative fiscal staff and one senator reported this as a view held by many legislators.

Question 13: Does sufficient evidence exist to judge the adequacy of how well assessment results are disseminated to intended audiences?

Major Evidence: Documents - 6, 7, 48

Interviews - 3, 4, 10, 16, 19, 21, 27

Panel Response: Evidence (e.g., survey results) was seen by panel members on the extent to which teachers and administrators receive data. However, no evidence was seen on the extent to which parents receive data. Also, interviews with legislative personnel indicated a lack of information reaching legislators. It is not clear, however, the extent to which legislators need test results; they may be more interested in how well others use the results. Improvements have been made in the speed with which results are returned to the users.

Question 14: Does sufficient evidence exist to judge the adequacy of the general process used for test development and the procedures used to revise tests?

Major Evidence: Documents - 1, 2, 17, 21, 35, 56, 57, 58, 61, 68

Interviews - 2, 16, 21, 26

Panel Response: Evidence exists that teachers, measurement specialists and content specialists are involved in the test development process, such as at the item writing

stage. Adequate evidence exists that item analyses are being done and systematically used in test revision. However, it would be useful to have a written statement of specifically how people and information are involved in the development process. No overall plan was seen for when items will be replaced and objectives changed. An advisory group proposed a systematic development process, but it is unclear as to how or if the document reached MEAP staff or was used.

A systematic development process needs to be accepted by MEAP with specific statements of how each step is to be conducted (e.g., who is involved, the nature of their involvement, the process used, and the resources required). Such documentation could greatly facilitate the work and provide a basis for judging the quality of the development process.

Question 15: Does sufficient evidence exist to judge the adequacy of how bias is defined, identified and/or eliminated?

Major Evidence: Documents - 1, 2, 28-32, 35

Interviews - 7, 16, 21, 23

Panel Response: Based on a review of MDE documents and on interviews with numerous MEAP staff members, bias appears not to have been defined by those responsible for either developing, implementing or reviewing MEAP test development activities.[7] MEAP does not appear to have a formal policy or position statement relative to cultural, sexual, age or other types of bias. Based on interviews with a few MEAP staff members, a tacit policy appears to exist which assumes that those individuals responsible for developing MEAP test items will monitor all items for bias.

Tests need to be carefully reviewed to eliminate biases against the aged, handicapped or either sex in terms of roles or other social factors. Documentation is needed that shows that a position on how to handle

[7] A booklet by the MacMillan Publishing Company, *Guidelines for Creating Positive Sexual and Racial Images on Educational Materials*, 1975, has been used as a general guide by the test item developers. The actual extent to which it has been used is unknown.

bias is established. It may be helpful to address the following questions relative to bias and MEAP test items in the development of a position.

1. What types of bias must be eliminated and how is each defined?
2. What definitions of bias are relevant for an examination of MEAP items?
3. What procedural safeguards should be implemented relative to the general issue of bias and assessment?

It is suggested that a committee of individuals who are sensitive to and knowledgeable about testing and bias in the test development process be established. This committee should have specific review/monitoring responsibilities regarding cultural and other biases.

Question 16: Does sufficient evidence exist to judge the adequacy of the process for development and review of content objectives?

Major Evidence: Documents - 1, 16, 17, 35, 42, 66

Interviews - 16, 24, 26, 28

Panel Response: No documentations were presented to the panel which clearly outlined a process used for development and review of content objectives. However, this responsibility is outside MEAP and thus was one in which the panel may have missed documents. Interviewees suggested, however, that the process is unclear. A document stating the specific procedures which are used would be helpful in not only evaluating the process but in clarifying to all parties concerned what the steps are, their purpose and who and how people (groups) are involved.

Question 17: Does sufficient evidence exist to judge the adequacy of the process used to gather input and select content objectives to be tested?

Major Evidence: Documents - 17, 23, 35, 42

Interviews - 21, 26, 28

Panel Response: The documents stating the content objectives do not describe the objective selection process. Interviews

with specialists indicate some confusion about who is responsible for the selection of the objectives used as the basis for the assessment tests. For instance, in one content area, a person in the curriculum division stated that the assessment staff determined which objectives were tested. The interviewee desired to know more about how the assessment staff performed the task. On the other hand, the assessment staff said such decisions were made by the curriculum division.

The present process for selecting objectives and assessing them needs considerably more documentation to determine whether it is an adequate procedure. For example, there are at least 200 mathematics objectives which are suggested as ones to master by fourth grade. It is not clear that the particular sample of 30 covered by MEAP is more defensible than some other sample of 30.

Question 18: Does sufficient evidence exist to judge the adequacy of the sampling procedure (e.g., the basis for the decision to use every-pupil-testing rather than sampling)?

Major Evidence: Documents - 1, 16, 35

Interviews - 2, 7, 16, 18, 21, 26

Panel Response: Documents adequately show how sampling was done to judge quality. However, the question of how appropriate the procedure is, given MEAP's goals and effects, is not documented adequately. The program staff has apparently talked much and received input from various sources on the issue of sampling. However, it does not appear that the input has been well integrated to develop an alternative proposal on how sampling could be done. Some individuals or groups are committed to every-pupil-testing, others strongly advocate matrix sampling. Others do not understand the basis for sampling or do not see the multiple purposes for sampling students, items, objectives, content areas or subareas to provide an overall scheme of testing.

The issue of sampling is intimately tied to the purposes of testing and the intended and unintended effects. It may be wise to gather the necessary information on intended and unintended effects and then meet with relevant groups, decide what purposes promote the positive effects and what purposes have negative side effects. Then the sampling could be

considered in relation to minimizing the negative side effects and maximizing the positive ones.

For example, there are at least 200 mathematics objectives which could conceivably be mastered by fourth grade. One could conceivably design a testing battery which would get a better sampling of the math objectives that would be useful for school and district planning purposes and would not produce the problems which the present strategy is producing. One could design a sampling survey which would test every pupil, but which would give a better picture of the total eudcational program.

V. MANAGEMENT

The panel considered 16 questions related to the management of MEAP. Issues ranged from internal and external communication to personnel topics such as staff development and affirmative action. Below are the questions concerning management, the major sources of evidence and the panel's comments.

Question 1:	Does sufficient evidence exist to judge the adequacy of short term planning, timelines and the planning process?
Major Evidence:	Documents - 11, 13, 14, 23, 79-81, 85, 99-101 Interviews - 7, 17
Panel Response:	Evidence suggests careful short range planning. Timeline charts are prepared in the first quarter of each calendar year which cover the following 15-month period. Each MEAP division integrates its schedule with this master time frame. A management-by-objective system is being developed whereby employees and the Department can project responsibilities, relationships and resources on a yearly basis to allow for comprehensive evaluations.

Question 2:	Does sufficient evidence exist to judge the adequacy of the documentation of procedures?
Major Evidence:	Documents - 15, 22, 99 Interviews - 16, 21
Panel Response:	Considerable documentation, such as in the technical reports, occurs concerning general procedures but the documents seldom give details of the processes used (e.g., number of teachers involved in review of tests, the sequence of reviews and who makes final decisions). It is difficult to tell if such lack of detail exists because the process is not clear or if it is to keep the reports concise. Evidence from interviews, however, suggest that the former may be the case in some stages of test development and decision making on technical issues.

To be useful, the documents appear to need more precision concerning (1) the actual procedures used during a given year and (2) suggestions for how to improve in a subsequent year. In making decisions about how to improve the documentation process the staff needs to look at the reasons for documentation. Some documentation is required to ensure that decisions are distinguished from mere suggestions or the raising of issues. As the staff grows and staff turnover occurs, such documentation is critical to maintaining continuity and consistency, or at least to ensure that deviations from a previous course of action are intentional.

The panel also found that informal documents were not always dated or authorship was not given.

Question 3: Does sufficient evidence exist to judge the adequacy of long term planning, timelines and the planning process?

Major Evidence: Documents - 11, 23, 82, 96

Interviews - 7, 16, 18, 21

Panel Response: It is unclear whether or not the long range plan of 1976 is official. No planning process is documented. As a mitigating circumstance, it can be noted that the moratorium on expansion has "frozen" the plan so that the objectives have not had to be updated.

Question 4: Does sufficient evidence exist to judge the timeliness of results reaching users?

Major Evidence: Documents - 81, 102

Interviews - 16, 17

Panel Response: Careful records of time in transit and processing are kept for each district's test scores. The MRC processing time (including delivery time to MRC and back to the district) is only three to four weeks--a considerable improvement over early efforts. One problem receiving continuing attention is time lag associated with district handling of materials. For the 1977-78 administration materials will be returned to the building rather than the superintendent (unless otherwise specified).

Question 5:	Does sufficient evidence exist to judge the adequacy of how contracting is done, e.g., (1) the process/ basis used to decide to do contracting? (2) the basis for selection of a contractor? (3) the process used to monitor contracting?
Major Evidence:	Documents - 13, 14, 59, 62, 63, 67 Interviews - 17, 19
Panel Response:	Contracting in the last few years has been done with the same agency, apparently based on a feeling of satisfaction with that agency's performance. However, the next RFP will be opened to competitive bidding. Such a move seems very appropriate to ensure that the highest quality and most cost-effective contracting is being done. Documents clearly state the responsibilities of the contractor. The role of writing RFP's and monitoring contracts is clearly assigned within the MEAP.
Question 6:	Does sufficient evidence exist to judge the adequacy of the allocation of time to various endeavors?
Major Evidence:	Interviews - 16, 17, 21, 26
Panel Response:	During their review, the panel observed no formal procedures for determining allocation of time among tasks beyond the fact that each individual has a major area of responsibility (e.g., test development or dissemination). Each person apparently works rather independently in setting his or her own daily activities. The panel has no specific comment in regard to this issue other than to say that such allocation of time should be analyzed to see that it is consistent with the priorities of the program as a whole.
Question 7:	Does sufficient evidence exist to judge the adequacy of how assessment fits in with other Department of Education activities?
Major Evidence:	Documents - 23, 24, 95 Interviews - 7, 16, 18, 21
Panel Response:	Evidence exists on the use of assessment data for allocation of funds for compensatory education. No overall

Department statement exists showing the link of MEAP to other programs. Outside audiences may be particularly interested in knowing the link between curriculum specialists, MEAP and the GESTALT[8] program.

Question 8: Does sufficient evidence exist to judge the adequacy of the division of responsibilities among staff?

Major Evidence: Document - 98

Interviews - 7, 16, 18, 21

Panel Response: No documents were seen which specified the responsibilities of individuals or service areas. For example, the division of responsibilities between MEAP and curriculum specialists seems unclear. Curriculum specialists have the responsibility to oversee development of the objectives which serve as the basis for the assessment tests and to assist school districts in curriculum changes based on assessment results. Documents stating the objectives and interviews with specialists indicate some confusion about who is responsible for the selection of the objectives used as the basis for the assessment tests. It is recommended that a written statement of responsibilities be prepared to clarify roles, especially between the programs.

Question 9: Does sufficient evidence exist to judge the adequacy of the attention to personnel issues, e.g., (1) composition of the total staff in areas of affirmative action, equal opportunity, technical skills? (2) employee formal and informal evaluation? (3) staff training and development? (4) human relations?

Major Evidence: Documents - 90, 93, 94, 95, 98

Interviews - 5, 7, 16, 17, 21

Panel Response: Evidence did not indicate an active concern for personnel evaluation. The composition of the staff indicates a concern for affirmative action, equal opportunity and a balance of technical skills. No formal staff training program exists. No assessment of competency levels and needs of individual employees occurs once people are hired.

---

[8] Getting Educational Specialists Thinking and Acting on Learning Theory

Question 10:	Does sufficient evidence exist to judge the adequacy of the attention to human relations and procedures used to review support services?
Major Evidence:	Interviews - 16, 21
Panel Response:	Limited evidence exists. Several questions need to be addressed by MEAP: Is the line between "professional" and "nonprofessional" or "supporting staff members" definitely drawn, or is there an appreciation of importance of all jobs and the function of each as related to the whole? Does the "chain of command" provide for "upward" and "lateral" communication on a par with "downward" communication? Does the Department have a system whereby the achievements and contributions of all staff members are recognized in both formal and informal ways (so as to build a spirit of teamwork and mutual respect among employees)?
Question 11:	Does sufficient evidence exist to judge the adequacy of the use of advisory groups?
Major Evidence:	Documents - 83, 84, 89, 97 Interviews - 2, 16, 21, 26
Panel Response:	Documents and interviews indicate extensive use of advisory groups. No documentation was seen that indicated that advisory groups had been asked for formal feedback (e.g., via questionnaires) on problems or concerns about the operation of advisory groups.
Question 12:	Does sufficient evidence exist to judge the adequacy of the external communication processes beyond data reporting?
Major Evidence:	Documents - 9, 24, 86-88, 91, 92, 95 Interviews - 1, 5, 9, 11, 12, 13, 16, 29
Panel Response:	A detailed communication plan for the Department's office of public affairs was not provided during the panel's visit. Should such a document not exist, the panel recommends that consideration be given to the development of such a planning-operations statement. In regards to MEAP, a special section of the plan could be devoted to supportive and cooperative activities between the public affairs staff and MEAP

dissemination staff, at least to avoid confusion over news releases and relations with news media, but more importantly, to ensure a positive, ongoing communication program.

The Department's communication effort appears to consist of news releases, personal contact between external audiences (news media, school officials, etc.) and Department staff; news "bulletins," memos and letters to various audiences and similar traditional activities and processes. A statewide magazine to priority audiences was discontinued in 1974 because of budget problems and a perceived lack of response from its readership. The initiation of a new periodical which would provide information about the Department's many successes (as detailed for the panel by the Superintendent of Public Instruction), its challenges and operation should be given careful consideration.

Question 13: Does sufficient evidence exist to judge the adequacy of the internal communication process and decision making procedure?

Major Evidence: Documents - 24, 49, 91

Interviews - 5, 9, 16, 18, 21

Panel Response: Evidence exists that an initial effort has been made to structure an ongoing communication program for internal and external audiences. For each audience, a number of goals and activities to achieve the goals are stated. However, the statement lacks specificity as regards responsibilities, publications, time schedules, etc. This statement should be the basis for a detailed plan of communication over a two or three year period. The statement, to be reviewed and updated at least annually, should contain (1) a definition of the target audiences; (2) the identification of knowledge and action goals deemed appropriate to accomplish with each audience; (3) a listing of various publications, audio-visuals and processes (meetings, workshops, etc.) to be used; and (4) the assignment of MEAP and other Department staff to specific responsibilities. Planning regarding when activities should be initiated, conducted, completed and evaluated should be done as well as an analysis of budget factors, as appropriate.

Formal, internal communication, such as an internal newsletter whereby individual staff members are kept informed of other units' activities, does not exist.

The panel recommends that the Department consider producing an internal newsletter (e.g., two pages every other week) which covers Department and various service unit activities, meetings, calendar of events, notations of personal achievements, new publications, new employees, retirements, employment or advancement opportunities, etc.

Question 14: Does sufficient evidence exist to judge the adequacy of the procedures for redress of grievances by those affected by assessment?

Major Evidence: Interviews - 1, 5, 9, 16

Panel Response: No documents appear to exist detailing exact procedures citizens or others can or must employ to present grievances. Staff members are aware of complaints "by those affected by assessment" but no formal procedures exist to deal with such grievances.

Question 15: Does sufficient evidence exist to judge the adequacy of the ongoing formative evaluation?

Major Evidence: Documents - 47, 48, 49, 56, 68

Interviews - 16, 21

Panel Response: No overall plan for formative evaluation exists. Some formative evaluation is occurring in regard to technical aspects (test content, reporting procedures, dissemination workshops). Little is being done in relationship to goals, effects, management and costs. In certain areas--such as obtaining feedback on workshops and gathering data on revisions of test materials--considerable useful information is being gathered in an efficient and timely fashion. The developers of materials appear to be using such information in product revisions. However, on some broader issues, the formative evaluation activities appear to be primarily reactive rather than based on a comprehensive plan of what data would be useful and for what purposes.

Since major decisions will be made in the future concerning program expansion and test revision, it is a critical time to sort out the key issues which need to be addressed and to develop a careful plan for gathering vital data for key decisions.

VI. INTENDED AND UNINTENDED EFFECTS

The purpose of this aspect of the panel review was to examine the extent, nature and quality of the evidence available regarding the effects which MEAP is having upon education in Michigan. Toward this end, studies were sought out which would indicate the degree to which MEAP was accomplishing its stated objectives and/or the degree to which it was affecting (either positively or negatively) other aspects of the educational system. Below are the questions which the panel investigated relative to intended and unintended effects. Major sources of evidence and the panel's comments are given for each question.

Questions 1 and 2: What evidence exists that MEAP has an adequate process for determining intended and unintended effects? What evidence exists about the kind of information collected and how well MEAP uses it?

Major Evidence: Documents - 47-52, 75-78

Interviews - 9, 21

Panel Response: The general conclusion is that the number of studies done was surprisingly few, given the importance and controversial nature of the program. Also, the studies that have been done were of limited value because of their superficial nature and because response bias was inadequately handled. One study relevant to MEAP effects was a survey of school principals (Steele, 1976). This survey produced much valuable information about the generally positive attitudes the principals had toward MEAP (at least for the 75 percent who responded), as well as information which suggests possible improvements in the MEAP procedures. The cost of that study is not available, but because it was a doctoral dissertation, one can assume such studies are within the means of the Department. Such evidence is valuable both as a basis for improving MEAP and for justifying its existence to legislative and other interested parties.

In spite of their potential importance and relatively small costs, only two surveys seem to have been conducted by the MEAP staff itself. One was a Superintendents' Feedback Survey conducted in 1974

and the other was a 1976 survey of participants in MEAP workshops. The Superintendents' Survey obtained reactions to the shift from norm referenced to objective referenced tests, the provision of assistance to staffs for understanding MEAP, and the utilization of test results.

The primary conclusion from the Superintendents' Survey is that there seems to be less objection to the objective referenced test than there had been to the previously used norm referenced test. The other possibly useful finding was that the superintendents felt that the proposed word relationships test be included at local district option. This type of survey could be much more helpful if these questions were less superficial, smaller samples were drawn, and greater effort was made to establish specific ways in which MEAP can be useful at the district level.

One survey of participants in the 1976 workshop was restricted to those interested enough in MEAP to participate in its workshop. That survey did gather information useful in improving future workshops, and while that is important, it did not get at the problem of assessing the effectiveness of MEAP itself.

In terms of long-range effects, it appears that more explicit attention should be given to the nature of state level education decisions and how assessment can provide appropriate data for such statewide decisions. A major effect of MEAP to date is that the test results are being used to determine eligibility for receipt of about $22.5 million in compensatory education funds. No substantive law changes based on MEAP data were called to the attention of the panel. A senator noted that MEAP has not yet convinced his fellow senators that the data produced on grades 4 and 7 tests are worthwhile enough to justify expansion to other grades.

Question 3: What additional documented information is needed from the affected groups (see checklist) on effects (see checklist)?

Panel Response: A primary purpose of MEAP is to provide information useful in making educational decisions. Inadequate evidence exists for an analysis of the nature of those decisions and of the information that would make for more informed decisions. This is true of all levels of decision making--state, district, school and classroom.

The panel found a definite insufficiency of documentation related to effects. Many questions need to be asked. For example,

1. To what extent and specifically how do teachers, administrators, parents, legislators and others use the results?

2. What effects do test-taking have on children?

3. What uses are made of the test results which are not intended?

4. What detrimental uses have been made of data?

5. To what extent have stated goals been attained?

6. What proposed changes in assessment procedures are most acceptable to various audiences? (Expected effects)

7. What differences exist in terms of immediate and long term effects?

8. How have people's opinions changed about the educational process as a result of receiving assessment data?

9. Has public confidence and interest in education increased as a result of having access to data?

10. What effect do assessment results have on parents?

11. How have assessment data been used to make resource allocations?

12. What decisions have been made by principals, superintendents, school boards, state officials, legislators and others at the state, district, school and classroom levels, based on assessment results?

13. What effect has assessment data or the existence of an assessment program had on legislative actions?

14. Could the decisions made based on assessment data been made as well with other already existing information?

15. What effect has assessment had on school and district testing programs? (see Brzezinski, 1977)

16. Has MEAP stimulated interest in and application of criterion referenced testing as part of the instructional program?

17. Has the curriculum been distorted as a result of the MEAP encouraging schools to focus upon math and reading, or have these subject areas been distorted by focusing only on a subset of objectives?

18. What curriculum changes have been made as a result of deficiences documented in assessment results?

19. Have assessment results followed by curriculum revisions affected pupil performance?

VII. COSTS

The panel examined the question of whether adequate evidence exists concerning dollar costs and time costs of the assessment program as well as whether the cost information which is available has utility to the users. Below are the questions the panel investigated relative to costs. Major sources of evidence and the panel's comments are also given.

Questions 1 and 2: Does sufficient evidence exist to judge the adequacy of the dollar costs for the total program, including state and district level costs? Does sufficient evidence exist to judge the adequacy of the utility of the costs information?

Major Evidence: Documents - 69-74

Interviews - 3, 4, 7, 9, 17, 18, 21, 26

Panel Response: There appears to be no formal system for judging the utility of cost information nor any procedure for examining the time costs of assessment. In the former case the panel assumes that the financial reporting system meets Michigan accounting standards or, if not, will be subject to audit exemption. The panel did not attempt to review the financial aspects of the program at that level.

The panel did note that the operational aspects of cost control appear to be sound. The budget is detailed as to line items, organization sub-units and amount spent on external contracts.

A substantial improvement included in the 1977-78 RFP is detailed on costs of the many separate "consumable" items in the contract such as score sheets and test booklets. Such information will allow MEAP to compare MRC costs with other vendors.

The major cost problem appears to be the lack of long range financial planning which is undoubtedly linked to the difficulty of getting state board and legislative approval for program expansion. Nevertheless, MEAP appropriations have grown from $219,000 in 1970 to a high of $893,340 in 1975 and in the current year dropped to $868,300. MEAP's request for 1977-78 included a substantial increase to $1,451,300 which would add three positions, allow expansion to testing for kindergarten and grades 1 and 10 and

begin test development for grade 12. In the process of considering the request in the legislature, the amount was reduced to $1,080,000 which would be solely for every-pupil-testing at grade 10. (Appropriation decisions were about to be made at the time the panel conducted interviews.)

Two points are important in this context. First, MEAP does not have documented an overall projection of costs for a "complete" assessment program, and second, MEAP does not appear to have validated detailed information on the incremental costs of adding specific elements.

Question 3: Does sufficient evidence exist to judge the adequacy of the time costs to students, teachers and administrators?

Major Evidence: Document - 73

Interviews - 3, 4, 7, 9, 17, 18, 21, 26

Panel Response: The question of opportunity or time costs has not been explicitly addressed by MEAP except to the extent that it has limited the length of the general tests so as to reduce the amount of time spent on administration. The panel was not shown any documentation relative to the overall time spent on "testing" versus "instruction" in the Michigan schools. The panel surmises that MEAP comprises only a small share of overall testing time even in the two grades now having every-pupil-testing, but suggests that Michigan would be well served by an overall review and the development of a strategy for testing.

Somewhat related is the issue of "cost sharing" by the state and districts. Districts cannot receive supplementary services from Westinghouse-MRC on a cost reimbursable basis. The MEAP staff has had preliminary discussions of a "cost sharing" procedure which would allow the Department to offer expanded services to districts, but has not yet had an opportunity to explore the feasibility of such a program nor its attractiveness to the local districts. A review of overall testing needs conceivably could lead to a realignment of testing services provided by the local, state (or even national) levels.

In summary, the panel finds that MEAP possesses generally adequate procedures for monitoring operational financial costs but has not yet developed adequate long range financial planning nor a strategy for dealing with financial and non-financial costs on a coordinated state-local basis.

VIII. SUMMARY

This report presents the findings of an external advisory panel which reviewed existing MEAP documents and conducted limited interviews. The panel's purpose was to determine the extent to which MEAP has sufficient documentation to make judgments about its (1) goals and purpose, (b) technical aspects, (3) management, (4) intended and unintended effects and (5) costs. The report also presents occasional evaluative comments and recommendations regarding each of the five areas which may help MEAP staff members as they plan a formative evaluation strategy.

After reading the findings of the panel presented in Sections III-VII, the MEAP staff is encouraged to refer back to Section II. In Section II the panel suggests possible ways MEAP could respond to the panel's findings. At the end of Section II the panel members present eight decisions which they consider most important for MEAP to make. The findings in Sections III-VII present the concerns the panel has in regard to these decisions and suggests where additional information is needed to reach such decisions.

APPENDIX A

Panel Selection and Biases

Documentation of Evaluation Team Selection Procedures

The evaluation team members were selected by the following procedures:

Jack Schmidt, Loretta Shepard and Judith Moyer met in Denver to discuss the purposes of the evaluation and the qualifications which team members should have to be able to evaluate a large scale assessment program. Since an assessment program has social policy, technical and management aspects, these types of knowledge were desired in the evaluation of team members.

Nominations for evaluators were obtained from Dr. David Donovan, Dr. Edward Roeber, Dr. Jack Schmidt and Dr. Loretta Shepard. The nominees were then considered in terms of availability and the complementary skills of the team members.

The team was comprised of five individuals, all of whom are considered to be objective evaluators, each with at least two skills which would contribute to a credible, total evaluation. The team members have skills in areas of assessment in general; dissemination; test construction; test bias; evaluation; politics; social policy; and education in general. None of the individuals has been directly involved with the Michigan Educational Assessment Program even in an advisory capacity and none is presently directly employed by a state assessment program; however, most have experience working with state assessment programs. This lack of representatives of state assessment programs seemed to further contribute to the objectivity of the evaluation, because those who are employed by assessment programs could be expected to have a definite positive bias toward state assessment.

Preconceptions About Testing and Assessment

At the beginning of the April 27 meeting each panel member responded to a series of questions, some structured, some open-ended. Their responses are summarized below.

This questionnaire covers several general issues about testing and assessment. The formal questions, to which you should respond on the strongly agree - strongly disagree continuum, are intended as stimulis. The questions are intended to identify some of the important issues which you should consider in characterizing your position. However, structured questions cannot possibly capture all of the nuances in your beliefs and values so it is important that you write in a statement of your point of view. If you feel that you are unfamiliar with a topic or do not have an opinion, please indicate this in the open-ended section.

Criterion referenced tests vs. norm referenced tests	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Norm referenced tests typically do not measure what is taught in school.	-	-	2	3	-
Criterion referenced tests are difficult to interpret because there are few absolute criteria.	-	-	1	4	-
Criterion referenced tests generally assume better coverage of a content domain.	-	4	1	-	-
Norm referenced tests provide more interpretable results about whether obtained scores are good or bad.	-	1	3	1	-

Please state your position regarding the pros and cons of norm referenced and criterion referenced testing.

1. Norm referenced or criterion referenced testing may or may not be appropriate depending on testing objectives. If the concern is to determine placement or relative standing--in a way that is not constraining or prohibitive--then norm referenced testing is appropriate. If the concern is to determine what an individual knows with respect to specified objectives, then criterion referenced testing is appropriate.

2. Not familiar with details of norm referenced and criterion referenced testing.

3. Both have value for different purposes. I am not a testing expert but assume that criterion referenced testing is most useful for the use of schools in determining in which areas their students do well or poorly. For broader judgments some "norms" need to be established even if not the traditional form of norm referenced testing.

4. No comment.

5. I believe norm referenced and criterion referenced tests both have their place. The critical issue is sorting out the purpose for testing in a given situation and then selecting the kind of test which best fits that purpose. The nature of the objectives covered, the way the test was developed (basis for item selection) and interpretability of results in terms of test purpose are essential to consider. I also prefer using the term objective referenced rather than criterion referenced, or considering them as two separate kinds of tests. I am presently quite skeptical about some criterion referenced tests because of the lack of procedures for establishing criteria.

Resource allocation on the basis of test scores	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Federal and state resources should be allocated to local school districts according to the number of students who qualify by a poverty criterion.	-	4	1	-	-
. . .according to the number of students who qualify by a measure of educational deficit.	1	2	1	1	-
. , .according to the amount of gain in pupil achievement demonstrated by a program.	-	1	2	1	1

Please state your position regarding the best criterion to be used in allocating resources:

1. Probably the best criterion to use in allocating resources is one which relates most strongly to the identification of specific needs or educational deficits. Consequently my position favors the use of needs assessment measures which identify specific academic deficits as a means of allocating resources.

2. Best criterion must be based on local financial needs (as related to local effort) in light of educational needs; educational deficit, povery (causation) as opposed to "political" aid which appears often as a criterion for allocating resources.

3. I have formed no position on this and am doubtful about the value of picking one characteristic as the universal determinant for resource allocation. Students needing extra help and therefore extra resources should be identifiable by a variety of ways. A greater problem is the total size of resource pool which is usually less than needed to meet the needs of the typical student.

4. SES of school attendance area.

5. I believe funds should be allocated according to needs but not exclusively as measured by the second and third questions above. I'm concerned about not only the basis for funds but the purposes for which they are to be used. I would prefer some assessment of the needs in terms of programs, management, etc., included in allocation. I can only accept the third point if consideration is given to intervening variables.

General consequences of testing	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The negative side effects of testing are so great that it ought to be banned from the public schools.	-	-	-	3	2
The information gained from testing is sufficiently important to offset the costs in dollars, time and anxiety felt by some participants.	-	4	1	-	-
Testing is likely to be harmful unless it is used only by classroom teachers for diagnostic purposes.[1]	-	-	-	3	1
Testing is essential to ensure accountability in education.	1	3	1	-	-

[1] Only four responses here because one wise guy changed the statement. Consequently his response couldn't be tabulated. Oh, the hazards of attempting to structure people's responses!

Please state your position regarding the relative benefits and hazards of testing:

1. The relative benefits and hazards of testing vary depending on the overall purpose(s) of testing and specific and general ways in which testing is carried out.

2. Benefits: Evidence of Performance; Planning; Refinement of Method or Curriculum.

   Hazards/Problems: Agreeing on Criteria; Assigning Responsibility; Self-Interest (Restraints)

3. Tests are useful to the degree that they have been carefully validated (and administered) and to the extent that the audience understands what the test results indicate and do not indicate. There is obviously great opportunity for error but I see no alternative way to get the necessary information.

4. Testing, in some form, is essential to school operations.

5. I believe testing has a very rightful place in education to give an objective basis for decision making and checking on progress. I think it is often over used and misused--done without a purpose. I prefer emphasis on diagnostic testing but definitely feel there are other purposes for testing. I think testing can definitely be used in accountability but I don't think it can or need be the only critical factor in an accountability system.

Definition of test bias	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Tests of academic achievement are inherently biased against members of minority cultural groups.	-	-	1	4	1
Test bias is demonstrated whenever average performance is not the same for all minority groups.	1	1	3	2	-
Test bias is a problem with test format whereby some children who have a skill are unable to demonstrate it because of the way in which a question is asked.	2	2	-	1	-

Please state your definition of test bias and the kind of evidence you would seek to determine if a test is biased:

1. Test bias would seem to exist in cases where test results indicate a strong performance-competence relationship for certain groups of children, and a relatively weak performance-competence relationship for other groups of children.

   To determine if a test is biased I would consider: whether test performance differed among groups of children in terms of race, ethnicity and social status; whether test norms were appropriate for the groups tested; whether test items seemed "closed" or "open" with respect to cultural pluralism.

2. No comment.

3. A test is biased when it does not measure what it purports to measure. Correction requires clarity of objectives for the test items and statistical analysis of results on one item compared with others which give different insights on the same skills being measured.

4. I tend to focus upon item content and opportunity to learn content.

5. I would agree primarily with the last definition under "definition of test bias." I think we do a great disservice when adjusting test items so that all minority groups score the same on each item. We then destroy the possibility of uncovering the skills which various groups have been somehow deprived of learning. The danger in testing is when we do not fairly allow students to demonstrate their skills. It is difficult, however, to tease out the difference between the skill and the operational definition (i.e., item) used.

APPENDIX B

Memorandum of Agreement

Memorandum of Agreement

1. Charge

The external evaluation panel consisting of Dr. Beverly Anderson (leader), Dr. William Cooley, Dr. Albert Holliday, Dr. William Mosley and Dr. Augustus Turnbull has been invited by the Michigan Educational Assessment Program to: (1) review the objectives of the program, (2) evaluate the program in terms of the objectives, (3) provide recommendations for program improvement, and (4) use and evaluate the Shepard checklist for the evaluation.

2. Audiences (in priority order)

- Director of Research, Evaluation and Assessment Services (REAS), Michigan Department of Education
- Michigan Educational Assessment Program (MEAP) staff
- Michigan State Board of Education
- Consumers (educators, parents, public)
- The media
- Technical persons (especially in the area of evaluation of large scale assessment programs)

3. Report/Editing

The panel will be solely in charge of developing and editing its final report. MEAP staff may write and disseminate any separate statement (such as an endorsement, a rebuttal, a commentary or a descriptive piece). It is understood that the panel's report is to be as short and direct as possible and to be designed to communicate with the audiences designated for the report.

4. Dissemination

The external panel shall release its report to the Director of REAS following the completion of the report. The Director of REAS may choose to endorse or not endorse the report depending on judgment of the quality and appropriateness of the report. Should the Director decide to disseminate a document describing the report, it will be identified as different than that of the committee. Only the committee's final report as edited by the committee will be distributed with the names of the committee on it.

In addition, the release of published information from the report shall have prior approval by the Director of REAS, except aspects of the report which deal with the evaluation of the Shepard Checklist.

5. Format of the Report

The following items were identified as desirable ingredients for the panel's final report:

a. citation of the memorandum of agreement
b. presentation of the major findings and recommendations
c. presentation of minority opinions, if any.

6. Questions to be Addressed in the Report

Specific questions to be addressed will include at least:

a. the panel's recommendations for change and further study

b. technical standards of tests

c. appropriate and inappropriate use of test results

d. merit of the objectives on which Michigan assessment is based

e. appropriate standards and procedures in test development

f. planning and development efforts

g. cost benefit projections for the program

h. usefulness and acceptance of assessment outcomes and reports for different audiences in Michigan

i. problems of bias in the Michigan Educational Assessment

j. administration procedures

k. staff organization and adequacy

7. Resources (budget) to Support the Program

Sufficient resources will be made available to the external review panel to support ten days of work by the team leader and five days of work for each of the other panelists, whatever secretarial support is needed in conducting the program, and whatever materials and equipment requirements, for example, tape recorders, taping, etc., in the Lansing site visit. It is understood that if any of the panelists need to make long distance telephone calls to collect opinions about the program from people in Michigan they may use the Michigan Department of Education telephone system or otherwise bill telephone calls to the Michigan Department of Education.

8. Delivery Schedule

The panel is to deliver its final report by June 16.

9. Access to Data

It is understood that the Michigan Department of Education will make available to the panel any and all data and reports required by the panel to do the job. This, of course, is restricted to those data and reports that are now available to the Michigan Department of Education.

10. Procedures

Pursuant to the above conditions the external five-person panel will have control over the evaluation process that it must implement to responsibly respond to the charge to which it has agreed. In accordance with this position the panel has agreed to implement the Shepard Checklist for evaluation of assessment program. The Michigan Department of Education has agreed to provide the training in the use of the Shepard Checklist to all panel members.

APPENDIX C

Shepard Checklist and Resulting Questions

Modified Shepard's Checklist for Evaluating Assessment[1]

A. Goals and Purposes

Kinds of Goals:

- Pupil diagnosis
- Pupil certification
- Program evaluation
  - program improvement
  - program selection
- Resource allocation
- Accountability
- Research
- Informing the public

Criteria for Judging Goals:

- Importance
- Uniqueness
- Feasibility
- Scope

Process for Goals Development

B. Technical Aspects

Tests:

- Criterion-referenced vs. norm-referenced
- Content validity
- Cultural bias
- Field testing *and revision*
- *Development procedures*

Sampling

Administration

Data analysis:

- *For test revision*
- *For reporting*

Reporting:

- (Data analysis)[2]
- Different reports for different audiences
- Interpretations
- *Credibility*
- *Pervasiveness (dissemination)*

C. Management

Planning

Documentation of process

Formative evaluation

Personnel

Equal opportunity

Support Services

---

[1] Items in italics were added based on the Michigan experience.

[2] Data analysis was made a more major category.

Subcontractors

Effective decision-making procedures *and communication mechanisms*

Redress of grievances

Timelines

Division of Responsibilities

Time allocations

Relationship to other programs

Advisory groups

D. Intended and Unintended Effects

People and Groups Who May Be Affected:

- Students
- Teachers
- Parents
- Principals
- District superintendents
- District curriculum specialists
- Regional personnel
- State curriculum experts
- State legislators
- Taxpayers
- Employers
- Special interest groups
- Researchers
- *Administration organizations*
- *Key communicators at local and other levels*
- *Curriculum organizations*
- *Other state departments of education*
- *News media*
- *Teacher organizations*
- *PTA's*
- *Federal legislators and legislators in other states*

Kinds of Effects:

- Outcome and process effects
- Immediate and long-term effects
- *Direct and indirect effects*
- *Knowledge*
- Decisions
- Laws
- Resource allocation
- Effects of assessment technology
- Opinions

Criteria for Judging Effects:

- *Need*
- *Market*
- *Consistency with goals*
- *Appropriateness of use*

E. Costs

Dollars

Time

Specific Questions Derived from Shepard Checklist for Use with MEAP[1]

A. GOALS AND PURPOSES

Does sufficient evidence exist to judge the adequacy of . . .

1. the established goals?
2. consideration given to the following kinds of goals:
   a. pupil diagnosis
   b. pupil certification
   c. program evaluation, improvement, selection
   d. resource allocation
   e. accountability
   f. research
   g. informing the public
3. the importance, uniqueness, feasibility and scope of the goals?
4. *the process used to establish goals?*[2]
5. *the existing process to modify goals?*
6. *the clarity of understanding among affected groups concerning what the goals are?*

---

[1] The questions are stated such that the adequacy of evidence on each topic could be assessed. Had the task been to actually judge the adequacy of each aspect of MEAP, the lead part of the questions could have been changed to read "Does the evidence indicate adequacy of . . ."

[2] Questions in italics correspond to italicized items on the checklist.

B. TECHNICAL ASPECTS

Does sufficient evidence exist to judge the adequacy of . . .

1. the basis for selecting norm-referenced or objective-referenced tests?
2. the consistency of goals to the choice of norm-referenced or objective-referenced tests?
3. the process used to select content areas for testing?
4. the process for development and review of content objectives?
5. the process used to gather input on objectives to be covered?
6. the process used to select content objectives to be tested?
7. the match of test items to content objectives?
8. how cultural bias is defined, identified, and/or eliminated?
9. the procedures used to field test items/tests?
10. *the procedures used to revise tests?*
11. *the general process used for test development?*
12. the sampling procedure?
13. the test administration process and method of handling missing data?
14. *the conduct and use of item analysis?*
15. longitudinal analyses?
16. the usefulness of reporting variables?
17. reporting procedures (e.g., appropriateness of data analyses, use of different reports for different audiences, appropriateness of interpretation)?
18. *how credible audiences consider the assessment data to be?*
19. *how well results are disseminated to intended audiences?*

C. MANAGEMENT

Does sufficient evidence exist to judge the adequacy of . . .

1. long and short term planning (e.g., use of PERT or other management schemes)?
2. timelines?
3. the planning process?
4. documentation of procedures?
5. ongoing formative evaluation?
6. attention to personnel issues, e.g.,
   a. composition of the total staff in areas of affirmative action, equal opportunity, technical skills?
   b. employee formal and informal evaluation?
   c. staff training and development?
   d. human relations?
7. procedures used to review support services (e.g., supplies, secretarial assistance)?
8. how subcontracting is done, e.g.,
   a. the process/basis used to decide to do subcontracting?
   b. the basis for selection of a subcontractor?
   c. the processes used to monitor subcontracting?
9. the internal communication process and decision-making procedures, e.g.,
   a. *how elements of the agency communicate?*
   b. how managerial and policy decisions are distinguished?
10. procedures for redress of grievances by those affected by assessment?
11. the timelines of results reaching users?
12. *the division of responsibilities among staff (both within the assessment staff and between assessment staff and other groups working on tasks impinging on assessment)?*
13. *the allocation of time to various endeavors?*
14. *how assessment fits in with other state department of education activities?*
15. *use of advisory groups?*

D. INTENDED AND UNINTENDED EFFECTS

1. What additional documented information is needed from the affected groups (see Checklist) on effects (see Checklist)?
2. What evidence exists that MEAP has an adequate process for determining intended and unintended effects?
3. What evidence exists about the kind of information collected and how well MEAP uses it?
4. Does evidence exist on the following effects?[3]
   a. To what extent and specifically how do teachers, administrators, parents, legislators and others use the results?
   b. What effects does the test taking have on children?
   c. What uses are made of the test results which are not intended?
   d. How has the assessment affected the curriculum of those areas assessed at local and state levels?
   e. How has assessment affected the curriculum in areas not tested at local and state levels?
   f. How has assessment affected teaching at grade levels other than the ones at which testing occurs?
   g. What effect has assessment had on school and district testing programs?
   h. To what extent have stated goals been attained?
   i. What differences exist in terms of immediate and long term effects?
   j. How have people's opinions changed about the educational process as a result of receiving assessment data?
   k. What effect has assessment data on the existence of an assessment program had on legislative actions?
   l. What detrimental uses have been made of data?
   m. What detrimental uses do people fear will be made of data?
   n. How has assessment data been used to make resource allocations?
   o. Could the decisions made based on assessment data been made as well with other already existing information?
   p. Has public confidence in education increased as a result of having access to data?
   q. Has the public become more interested in education as a result of having access to data?
   r. What proposed changes in assessment procedures are most acceptable to various audiences? (Expected effects)

---

[3] The questions listed in this category are ones explicated after the planning meeting and appear especially important for Michigan to investigate.

E. COSTS

Does sufficient evidence exist to judge the adequacy of . . .

1. the dollar costs for the total program, including state and district level costs?

2. the time costs to students, teachers and administrators?

3. the utility of the cost information?

APPENDIX D

A Review of the Michigan Educational Assessment Program, March 1977

A REVIEW OF THE MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM (MEAP)

Accountability

The Michigan accountability model is actually a planning model for rational decision making. Each step of the model follows logically from the preceding step, and each generates the activities for the next step. The six steps of the model are:

1) The Identification of Common Goals
2) The Development of Performance Objectives
*3) Assessment of Needs
4) Analysis of Delivery Systems
5) Evaluation
6) Recommendations for Improvement

The assessment program is a part of the six-step accountability model developed by the Michigan Department of Education. Though the assessment program began prior to the completion of the first two steps of the model, it is now an integral part of the accountability process.

The accountability concept is based on the assumption that informed people will make better decisions. Accountability requires that all participants in the educational process become informed about how the system is working and use that information to improve the system.

Educational accountability requires the gathering and reporting of information on what educators are trying to accomplish, how well they are doing, what methods are being used with what success and at what costs, and what changes are recommended. Accountability can be used by local educators at the building and district level or at the state level to evaluate the overall status and progress of the state's educational efforts. Only by periodically gathering systematic information can areas of weakness be improved and areas of strength maintained.

History

The Michigan Educational Assessment Program (MEAP) since January, 1970, has provided information which contributes to the assessment of educational needs. The Program was initiated by the State Board of Education, supported by the Governor, and funded by the Legislature initially through enactment of Act 307 of the Public Acts of 1969, and subsequently under Act 38 of the Public Acts of 1970.

Norm-Referenced Tests

The first four years of the assessment program (1969-1973) made use of standardized, norm-referenced testing instruments to assess student performance in vocabulary, reading comprehension, English usage, and arithmetic. In addition, attitudinal and socioeconomic data were gathered in certain years.

Prepared by Martha S. Caswell of the Michigan Department of Education, March 1977.

However, it was decided that the information provided by the norm-referenced tests did not adequately serve the purpose of MEAP to provide information to the State Board, the Executive Office, the Legislature, local educators, teachers, students, and parents on the status and progress of Michigan basic skills education. Therefore, the norm-referenced tests were replaced by objective-referenced tests in the 1973-74 school year.

Objective-Referenced Tests

In accordance with Michigan's educational goals, a series of learning areas were selected for assessment at five grade levels (1, 4, 7, 10, and 12):

Preprimary (objectives for entry to Kindergarten and grade 1)

Art
Communication Skills
Foreign Languages
Health Education
Mathematics and Metric Measurement
Music
Physical Education
Science
Social Studies

Minimum objectives for end of grades 3, 6, and 9)

Life Role Competencies (Objectives for end of Grade 12)

Objectives to be Tested

The minimal skill performance objectives for these areas were developed under the direction of staff from the Instructional Specialist Program of the General Education Services, and with the help of committees of Michigan educators and citizens. The State Board received and approved each set of objectives (except for the Life Role Competencies which are still in a developmental stage). Each set will be reviewed and revised periodically. These performance objectives do not cover all skills in each subject area but are, instead, minimal skill objectives which most students are expected to attain. Local school districts should supplement these objectives with others in order to provide pupils with the breadth and depth of educational experiences needed to achieve their potential.

Grades 1, 4, 7, 10, and 12 were selected as the five points at which to assess student performance because they represent important transition points in each student's education: the beginning of primary education, completion of primary education, completion of middle school education, and the completion of high school.

The Grade 12 Life Role Competencies are in a developmental stage at this point in time. The idea of using them as the basis for some kind of 12th grade exit test has been discussed, but as of now, no plan has been developed for assessing these objectives. How the objectives will be used depends in part on the characteristics of the objectives themselves.

The other selected subject areas tested in 1976-77 were music and metric measurement. In addition to providing statistical data on the behavior of test items, these tests also give an indication of how students are doing on a statewide basis in the subject areas selected. The two areas tested in previous years are health and science.

Grade 10

1976-77 was the second year that testing has been done on the tenth grade level on a voluntary basis. Some 170 schools volunteered to participate in the program this year. Selected objectives were tested on an every-pupil basis in the volunteer schools in the subject areas of reading and mathematics. There were four different test forms, two for reading and two for mathematics. Each objective was tested by a set of eight items, five items for reporting and three for item validation. In general, the reaction of participating schools to the tenth grade program has been positive.

Kindergarten

The Kindergarten testing program is also conducted on a volunteer basis. The Kindergarten program has been offered for 2 years now and contains several feature that are different from the testing program at Grades 1, 4, 7 and 10. Testing children in this age group requires the use of testing methods that differ from those used with older children (i.e., paper-and-pencil items).

The Kindergarten tests are based on the Preprimary Performance Objectives. Teachers are provided with the objectives, suggestions for ways of testing the objectives, and forms for recording student performance. Testing is done throughout the school year. Teachers may test any or all of the objectives and can use either the tests MEAP provides, their own tests, or a combination of the two. The Kindergarten assessment tests are provided as a service for the teachers and they are encouraged to use them in the way that will be most useful for them. Meetings are held periodically with Kindergarten teachers throughout the state to give teachers the opportunity to provide MEAP with feedback and suggestions for test improvement.

Grade 1

The first grade assessment program was introduced in the 1974-75 school year. Even though it is now three years old, it is still considered a pilot program. Again, because of the problems involved with administering a paper-and-pencil test to this age group, tests are administered on an individual basis or in a small group situation. This involves a good deal of teacher time so the Preprimary Objectives selected to be tested are few in number and the number of items per objective is also limited.

The tests are given on a sampling basis so that not all schools or first graders in Michigan are involved. The test results for the statewide sample yield reliable data on the objectives tested. This information is used to improve test items and administration procedures, for revising the objectives themselves, and for determining the extent and nature of educational needs as reflected by the results.

Test Reports

There are a number of test result reports for grades 4, 7, and 10, which are returned to teachers, principals, and district personnel. These report forms contain various kinds of information which are of interest to educators at different levels. The report forms provided are the Individual Student Report, the Classroom Listing Report, School Summary Report, District Summary Report, the Test Item Analysis, and the Proportions Report.*

A new feature of the Assessment Program this year was the Proportions Report provided to districts as part of the fourth and seventh grade test results. This Report indicates for a district and the schools within a district the percentage of students attaining each of four proportions of the mathematics or reading objectives.

Future Directions

Using the Proportions Report and three criteria, the state will identify schools with high, moderate, or low needs in reading and/or mathematics. Once high needs are identified, the state will work with a school which asks for help to ensure that students master the basic skills. This program is in the developmental stages at this point in time. The Department hopes to operationalize it in a few selected schools on a trial basis in June, 1977.

Pending action by the State Board of Education, tenth grade testing will be done on an every-pupil basis in reading and mathematics next year. The State Board should approve or disapprove this plan by April, 1977. If the Board approves it, the Assessment Program is ready to implement the program in the fall.

Next year Art and Social Studies tests will be administered on a sample basis. This will result in a total of six performance objective areas (other than reading and mathematics) which will have been tested on a statewide basis. After all seven subject areas (for which performance objectives have been developed) have been tested, the process will begin again.

The Assessment Program, because of recommendations from math specialists, will include several metric objectives as part of the grade 4 and 7 mathematics objectives tested next year. These metric objectives will be in addition to the regular math objectives tested each year. As the transition is made from the measurement system now used to the metric system, more metric objectives will be included on the assessment tests.

The Assessment Program, as part of the accountability model, will continue to assess the status and progress of Michigan basic skills education. The assessment data will be used to identify schools with low, moderate, and high needs in the areas of math and reading. The use of assessment data in this way will enable the state to work with schools to continue to improve the quality of education received by Michigan students.

---

*For those interested in a more detailed discussion of the report forms, see the Interpretive Manual, 1976-77, Michigan Department of Education, Lansing, Michigan 48909.

APPENDIX E

Documents Reviewed and People Interviewed

Documents Reviewed

Document references are not always complete but should be sufficient for identification. MEAP has a file of the listed documents. Nearly all documents were prepared by the Michigan Department of Education, Michigan Educational Assessment Program or a contractor of MEAP.

1. MEAP: A Technical Analysis of the Michigan Assessment of Basic Skills (Prepared by Michigan Association of Professors of Educational Administration, Task Force on Educational Assessment and Accountability) (no date)
2. The Michigan Educational Assessment Program: An Abstract (no date)
3. The Michigan Educational Assessment Program: Recommendations for Modification (no date)
4. House, E., Rivers, W., & Stufflebeam, D. An Assessment of the Michigan Accountability System. March 1974
5. MDE. A Staff Response to the Report: An Assessment of the Michigan Accountability System. May 1974
6. A Response to the Michigan Education Department's Defense of Their Accountability System. Stufflebeam, D. August 1974
7. Murphy, J. T., & Cohen, D. K. Accountability in Education--the Michigan Experience (no date)
8. A Position Statement on Educational Accountability. 1972
9. Questions and Answers About the Michigan Educational Assessment Program (no date)
10. Charting Educational Progress (no date)
11. Timeline - Art, Social Studies, First Grade, Kindergarten (no date)
12. Timelines, 1976-77 and 1977-78
13. Proposed Contractor/MDE Calendar for 76-77 MEAP
14. Proposed Contractor/MDE Calendar for 77-78 MEAP
15. MDE. 1975-76 Technical Report (4th Report)
16. 1972-73 Test Development and Experimental Tryout - Technical Report. 1973
17. Technical Report, Michigan Tenth Grade Project - Westinghouse Learning Corporation. October 1975
18. First Report - Objectives and Procedures 1974-75. August 1974

19. Second Report - Individual Student and Classroom Reports: Explanatory Materials 1974-75. October 1974

20. Third Report - School and District Reports: Explanatory Materials 1974-75. October 1974

21. Recommendations for Tested Objectives, Items, Booklets and Procedures, Phase I Report of the 1974-1975 Michigan Educational Assessment Test in Reading. April 1975

22. 1975-76 MEAP Objectives and Procedures

23. The Long Range Plan of the Michigan Educational Assessment Program. May 1976

24. A Review of the Michigan Educational Assessment Program. March 1977

25. MDE. 1976 Status Report on Michigan Basic Skills Education

26. MDE. 1976-77 MEAP Interpretive Manual

27. Michigan Department of Education Memorandum (4-14-77) Re: Plans for Gathering Nationally-Comparative Data in MEAP (MEAP Task Force Minutes Attached)

28. Michigan Educational Assessment Program 1974-75 (Year 6) - Student Assessment Booklet, Grade 7 (Copyright 1973, 1974)

29. Michigan Educational Assessment Program 1975-76 (Year 7) - Student Assessment Booklet, Grade 4 (Copyright 1973, 1974, 1975)

30. Michigan Educational Assessment Program 1975-76 (Year 7) - Student Assessment Booklet, Grade 7 (Copyright 1973, 1974, 1975)

31. Michigan Educational Assessment Program 1976-77 (Year 8) - Student Assessment Booklet, Grade 4 (Copyright 1973, 1974, 1975, 1976)

32. Michigan Educational Assessment Program 1976-77 (Year 8) - Student Assessment Booklet, Grade 7 (Copyright 1973, 1974, 1975, 1976)

33. MEAP 1976-77 Test Results - 4th or 7th Grade Test Results Folder for Teacher

34. MEAP 1976-77 Test Results - 4th or 7th Grade Test Results Folder for Principal

35. Development of Test Items and Instruments in Reading and Mathematics for the 1973-74 Michigan Educational Assessment Program. 1974

36. Where Research, Evaluation and Assessment Fit In. April 1973

37. MRA Phase I Report and Recommendations. 1975

38. MRA Phase II Report and Recommendations. 1975

39. MCTM MEAP Mathematics Interpretive Reports (3). 1970-73, 1973, 1974

40. The Michigan Educational Assessment Program: Selected Working Papers (Report Prepared by Michigan Association of Professors of Educational Administration, Task Force on Educational Assessment and Accountability) (no date)

41. "Between Us" - Editorial in "Michigan Elementary Principal." Total of eight columns printed from November/December 1970 to Winter 1976. Herbert C. Rudman, Michigan State University

42. Minimal Performance Objectives for Mathematics Education in Michigan. 1973

43. Ranking of Michigan Districts by Selected Financial Data 75-76

44. The Common Goals of Michigan Education. 1977

45. The Public's Understanding and Attitudes Toward Educational Accountability. 1974

46. Examples of Letters and Memos Regarding MEAP (especially to outside audiences)

47. Brzezinski, E. J. Testing Programs and Activites in Michigan Pubicc Schools, Fall 1976. (Working draft, April 1977)

48. Aquiro, A. Teacher Attitudes Regarding the Use of the 1974-75 Michigan Educational Assessment Program Test Results. July 1975

49. Steele, D. J. An Investigation of Administrative Practices and Attitudes Related to the Use of Michigan Educational Assessment Program Test Results. 1976

50. Michigan Department of Management and Budget Correspondence and MEAP Questionnaire, Milstein (May 11, 1977)

51. A Survey of Views Concerning Michigan 4th Grade Assessment Program 1973-74 (MEA)

52. Superintendent's Feedback Survey (MEAP). 1974

53. MRA's Review of the MEAP 10th Grade (Alexander). 1977

54. Record Center Activities (no date)

55. MEAP 76-77 Item Analysis. 1976

56. Items Selected to be Replacement Items, 1976-77 (no date)

57. Michigan Department of Education Memorandum (March 30, 1966) Re: Procedures for Item Replacement in MEAP Mathematics and Reading Tests

58. MEAP Memorandum (11-12-75) Re: Item Replacement Strategy for 1976-77 MEAP

59. RFP for Technical Support Services for the Development of Test Items and Instruments in Metric Education and Social Studies. October 1975

60. MEAP Fund Accounting (no date)

61. MEAP Memorandum (9-8-75) Re: 1976-77 MEAP (Fourth and Seventh Grade Core Program)

62. Michigan Department of Education Correspondence (8-15-74) Re: Request Detroit Public School District to participate in test development projects during 1974-75 school year

63. Michigan Department of Education Memorandum (4-1-77) Re: 1977-78 MEAP (Description of program/activities)

64. Untitled Outline (begins. . ."As a result of MEAP dissemination activities, the following audiences. . .") (no date)

65. Michigan Department of Education Correspondence (To Assessment Coordinators from Ed Roeber) (5-10-76) Re: Summary of Plans for 1976-77 MEAP

66. Minimal Performance Objectives for Math Education - A Metric Appendix (no date)

67. RFP to Provide Support Services for the Administration, Scoring and Reporting of Educational Assessment Instruments in the 1977-78 MEAP. May 1977

68. Michigan Department of Education Memorandum (2-23-77) Re: Needs Assessment Survey 1976-77, Summary of Data Results

69. MEAP Fund Accounting Exhibit B (no date)

70. Projected MEAP Budget by Major Category (no date)

71. Assessment: FY 76-77 Spending Plan by Unit (no date)

72. REAS Program Planning Document. 6/11/76

73. Roeber Memo on Shared Cost. 12/30/76

74. Annual Appropriations Totals 70-77 (no date)

75. Enrolled House Bill 3886 Act 38 of 1970

76. Chapter 3 State School Aid Act 1976-77

77. DOE/SBE/State Aid to Improve Achievement in Cognitive Skills (Rules) 4/25/72

78. Changes in Chapter 3 Administrative Rules 75-76, 76-77

79. Westinghouse Learning Corporation Proposal (no date)

80

80. Record Center Activities (no date)

81. Records of Time in Transition and Processing (no date)

82. Superintendent of Public Instruction Memorandum (8-26-75) transmitting 1975-76 CAAP to State Board of Education

83. MEA Recommendations re MEAP, April 2, 1976 and December 21, 1976

84. Three pages of references re MEA action re MEAP between Fall 1969 and Spring 1974, with appendices A-C

85. CAAP-MEAP. August 26, 1975

86. "In the News" report of newsclips, compiled by MDE Office of Public Affairs. April 18 and May 9, 1977

87. Michigan Education, Spring 1974 and Summer 1974

88. MDE news release regarding "4-year trend on state assessment tests." March 2, 1977. Memorandum from Dave Donovan to Tom Farrell. March 3, 1977

89. Memorandum regarding "meeting with two members of the MASB" from Judith Moyer to MEAP staff. August 12, 1976

90. Memorandum regarding "Telephone and receptionist responsibilities," from Dave Donovan to REAS staff. February 3, 1977

91. Nine-page reference to major audiences and dissemination activities, MEAP (no date)

92. Memorandum regarding "Receipt of the booklet Do You Use MEAP Test Results Appropriately" from John W. Porter to State Board of Education. April 21, 1977

93. Policy Statement regarding equal employment opportunity and affirmative action program, MDE. December 7, 1972

95. Probationary Service Rating. 1972

96. Memorandum regarding Approval of Rescinding of Moratorium on Expansion of MEAP to State Board of Education. March 24, 1977

97. Agendas regarding MEAP Task Force meetings and meetings of the REAS Advisory Council. March and April 1977

98. Memorandum regarding Orientation of New Staff, from Judith Moyer to MEAP staff. June 22, 1976

99. MDE memorandum (4/1/77) Re: 1977-78 MEAP (Description of program/ activities)

100. Three-year dissemination plans for MEAP (no date)

101. Memorandum on 1976-77 MEAP (Huyser/Fishbein). September 8, 1975

102. Olsen Memorandum on Turnaround. December 28, 1976

103. Status Report on Michigan Basic Skills Education, 1976-77 (no date)

People Interviewed

1. Ashmore, Mary Kay; Administrative Assistant for Public Affairs, Michigan Association of School Boards
2. Brictson, Paula; MEAP Research Consultant, MDE
3. Brown, Dan; Legislative Fiscal Agency Staff, Michigan House of Representatives
4. Carillot, Vince; Legislative Fiscal Agency Staff, Michigan House of Representatives
5. Caswell, Martha; Educational Technician, MEAP Dissemination Unit, MDE
6. Chung, K-Such; Research Consultant, MDE
7. Donovan, David; Director, Research, Evaluation and Assessment Services, MDE
8. Ebel, Robert; Professor of Education, Michigan State University
9. Farrell, Tom; Assistant Superintendent for Public Affairs, MDE
10. Faxon, Senator Jack; Chairman, Michigan Senate Education Committee
11. Firestone, Joanna; Capitol Correspondent, United Press International
12. Landis, Elwood; Editor, Teacher's Voice, Michigan Education Association
13. McDermott, Larry; Capitol Correspondent, Associated Press
14. Mehrens, William; Professor of Education, Michigan State University
15. Moore, Jack; MEAP Research Consultant, MDE
16. Moyer, Judith; Coordinator, MEAP Dissemination/Utilization and Analysis Unit, MDE
17. Olsen, June; Coordinator, MEAP Test Administration Unit, MDE
18. Porter, Jonn W; Superintendent of Public Instruction, State of Michigan
19. Rinehart, Paul; Education Analyst, Department of Management and Budget, State of Michigan
20. Rodenbaugh, Jill; MEAP Education Technician, MDE
21. Roeber, Edward; MEAP Supervisor, MDE
22. Rudman, Herbert; Professor of Education, Michigan State University

23. Schooley, Daniel; Coordinator, Compensatory Education Evaluation Unit, MDE

24. Scott, Wayne; Math Specialist, MDE

25. Sederburg, William; Michigan Legislative Assistant

26. Shakrani, Sharif; Coordinator, MEAP Test Development Unit, MDE

27. Sullinger, Gary; Legislative Fiscal Agency Staff, Michigan Senate

28. Trezise, Robert; Communication Skills Specialist, MDE

29. Wilkinson, Varl; Assistant Executive Secretary, Michigan Association of School Boards